0

U.S. Army Research Institute
for the Behavioral and Social Sciences

Research Report 1666

Using the Backward Transfer Paradigm to Validate the AH-64 Simulator Training Research Advanced Testbed for Aviation

John E. Stewart II
U.S. Army Research Institute

September 1994

Approved for public release; distribution is unlimited.

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Technical review by

John A. Dohme
David M. Johnson
Dennis C. Wightman

Accesion For		
NTIS CRA&I	X	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POX, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1994, September	Final Jan 93 - Oct 93

4. TITLE AND SUBTITLE
Using the Backward Transfer Paradigm to Validate the AH-64 Simulator Training Research Advanced Testbed for Aviation

5. FUNDING NUMBERS
63007A
794
2224

6. AUTHOR(S)
Stewart, John E., II

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-IR
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

8. PERFORMING ORGANIZATION REPORT NUMBER
ARI Research Report 1666

9. SPONSORING MONITORING AGENCY NAME(S) AND ADDRESS(ES)
--

10. SPONSORING MONITORING AGENCY REPORT NUMBER
--

11. SUPPLEMENTARY NOTES
--

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE
--

13. ABSTRACT *(Maximum 200 words)*
The Simulator Training Research Advanced Testbed for Aviation (STRATA) is a modular simulation system for the AH-64 helicopter. The Adams and McAbee (1961) backward transfer of training paradigm was used to determine if AH-64 piloting skills transfer to STRATA. Ten AH-64 pilots participated in the experiment. They were asked to fly a mission scenario consisting of 13 single-pilot Aircrew Training Manual (ATM) tasks. No orientation or practice sessions were given. Most participants rated STRATA as highly similar to the AH-64 in handling, but believed the visual display system had limitations for extremely low-altitude tasks like hovering. Performance ratings during the experiment by a senior instructor pilot (IP) indicated that of the total 130 task events, 88.5% were performed to ATM standards. After the experiment, four independent judges, all aviators, rank-ordered participant performance on the stationary hover task using output from STRATA's automated performance measurement system. Rankings showed agreement between judges and high correlation with IP ratings made during the experiment.

14. SUBJECT TERMS
Transfer of training
Flight simulation
Performance measurement
Helicopter simulation

15. NUMBER OF PAGES
41

16. PRICE CODE
--

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

Research Report 1666

Using the Backward Transfer Paradigm to Validate the AH-64 Simulator Training Research Advanced Testbed for Aviation

John E. Stewart II
U.S. Army Research Institute

Rotary-Wing Aviation Research Unit
Charles A. Gainer, Chief

Training Systems Research Division
Jack H. Hiller, Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

September 1994

Army Project Number
2Q263007A794

Education and Training

FOREWORD

The U.S. Army Research Institute for the Behavioral and Social Sciences Rotary-Wing Aviation Research Unit (RWARU) at Fort Rucker, Alabama, is committed to enhancing the readiness of Army aviation units through the development of effective training technology. The Simulator Training Research Advanced Testbed for Aviation (STRATA) is the cornerstone of this commitment. It was designed to support research to determine the training effectiveness of simulators and training devices. In its present configuration, STRATA represents the AH-64A helicopter. The research described in this report was initiated 25 January 1993 pursuant to the RWARU Research Task entitled Aviation Training Strategies for Improving Combat Readiness. The objective was to validate the current configuration of STRATA. This effort was internal to RWARU.

One simple method for determining the training effectiveness of a flight simulator is the backward transfer paradigm. Pilots highly experienced in the aircraft but unfamiliar with the simulator perform standard aviator tasks in the simulator without the benefit of prior practice. Successful performance of the tasks can be taken as evidence that the simulator is a valid representation of the aircraft.

Ten AH-64 aviators from an operational unit took part in the experiment. All flew the same mission scenario, which consisted of 13 generic aviator tasks from the Aircrew Training Manual (ATM).

Results showed that backward transfer did occur between the AH-64 and STRATA. Of 130 task events (13 X 10 participants), 88.5% were performed within ATM standards. Participants rated STRATA's handling characteristics as very similar to those of the AH-64A. This can be interpreted as evidence that STRATA is a valid simulation of the AH-64A.

These and other research findings from STRATA were briefed to the Deputy Commanding General, U.S. Army Aviation Center, Fort Rucker, Alabama, in August 1993 and to the Deputy Chief of Staff for Personnel, Department of the Army, Washington, D.C., in December 1993. The outcome of these briefings was an increased interest in STRATA as a tool for addressing critical training issues.

EDGAR M. JOHNSON
Director

ACKNOWLEDGMENTS

This experiment was successful because of the combined contributions of many people. Chief Warrant Officers (W-3) Al Eggerton and Mike Couch, both experienced AH-64 pilots, provided invaluable assistance with the mission scenario. Eggerton also evaluated the performance of experimental participants. Dale Weiler helped develop the premission briefings and played the role of air traffic controller. The intricacies of automated data capture were undertaken with aplomb by Rande Doty and Larry Murdock, who, along with Rolf Beutler, developed the control program that triggered recording events during the mission. A note of appreciation is also in order for Nick Donker, Fred Zalzal, and other members of the STRATA support team from CAE Electronics for their dedicated efforts.

USING THE BACKWARD TRANSFER PARADIGM TO VALIDATE THE AH-64 SIMULATOR TRAINING RESEARCH ADVANCED TESTBED FOR AVIATION

EXECUTIVE SUMMARY

Requirement:

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) developed the Simulator Training Research Advanced Testbed for Aviation (STRATA). The simulator was designed primarily for training effectiveness research for a variety of aviation training device configurations. While most conventional simulators are designed to support specific training objectives, STRATA is a true research testbed simulator purposely designed to allow for changes in hardware configurations. In its current configuration, it represents the AH-64A Apache helicopter.

Although flight simulators have become increasingly costly and complex, there is a paucity of empirical evidence as to how effective they are for training. Most often, the simulator is integrated into a training system with the assumption that piloting skills will transfer from simulator to aircraft and vice versa. This is an empirical question that can only be answered through transfer of training research. As a consequence, the validity of the simulator in terms of skills transfer is often unknown.

Procedure:

One relatively simple paradigm for determining the training effectiveness of a flight simulator is the backward transfer paradigm. Pilots highly experienced in the aircraft but unfamiliar with the simulator perform standard aviator tasks in the simulator without the benefit of prior practice. Successful performance of the tasks can be taken as evidence of backward transfer. If backward transfer is demonstrated, one can assume that forward transfer from simulator to aircraft would also occur.

Ten AH-64 aviators from an operational unit took part in the experiment. All flew the same mission scenario, which consisted of 13 generic aviator tasks from the Aircrew Training Manual (ATM). Examples of the tasks were stationary hover, hover taxi, straight and level flight, rolling takeoff, and single engine roll on landing.

Findings:

Results showed that backward transfer did occur between the AH-64 and STRATA. Of 130 task events (13 X 10 participants), 88.5% were performed within ATM standards. Participants rated STRATA's handling characteristics as very similar to those of the AH-64A. This can be interpreted as evidence that STRATA, as it is currently configured, is a valid simulation of the AH-64A.

Utilization of Findings:

The backward transfer results suggest that the present configuration of STRATA (e.g., fiber optic, helmet-mounted display, G-seat, AH-64 cockpit with full instrumentation) constitutes a valid training device for the sustainment of AH-64 piloting skills. The findings also suggest additional research using alternative configurations (e.g., a rear-projection visual display, no G-seat) in the same backward transfer paradigm to determine whether a simpler, less costly AH-64 simulator configuration would also provide a valid medium for skills maintenance. The same research paradigm can be applied to other aircraft simulators to determine cost and training-effectiveness tradeoffs in the design of simulators.

USING THE BACKWARD TRANSFER PARADIGM TO VALIDATE THE AH-64 SIMULATOR TRAINING RESEARCH ADVANCED TESTBED FOR AVIATION

CONTENTS

	Page
INTRODUCTION	1
Background	1
The Backward Transfer Paradigm	1
Previous Research	1
OVERVIEW OF RESEARCH APPROACH	2
Rationale	2
Objectives	2
METHOD	3
Participants	3
Procedure	4
DEPENDENT MEASURES	5
Self Reports and Subjective Measures of Performance	5
Automated Performance Measures	5
RESULTS	7
Participant Evaluation of STRATA	7
SIP Ratings of Performance	8
DRA Measures and Their Correlations	17
DISCUSSION	19
Participant Evaluations of the Simulation	19
Demonstration of Backward Transfer	19
Subjective and Objective Ratings and Rankings	19
Limitations	19
Suggestions for Future Research	20
REFERENCES	23
APPENDIX A. PARTICIPANT QUESTIONNAIRE	A-1
B. OPEN-ENDED COMMENTS FROM QUESTIONNAIRE	B-1

CONTENTS (Continued)

Page

LIST OF TABLES

Table 1. Participant Background and Experience 3

2. Performance Measures for ATM Tasks 6

3. Participant Ratings of Similarity of STRATA to Aircraft 8

4. Frequency Distribution of Performance Ratings for 13 Tasks 10

5. Frequency Distributions of Performance Ratings as a Function of Flight Experience 11

6. Pearson Intercorrelations: Flight Hours, Time Since Last Flight, and Checkride vs. UNSATS and Marginals 13

7. Ratings and Rank-Orderings of 10 Participants on Task 1: Pick Up to Hover 16

8. Performance Measures (Means and Standard Errors) for Initial Hover Task 17

9. Spearman Rank Order Correlations of DRA and Other Performance Measures for Initial Hover Task 18

LIST OF FIGURES

Figure 1. Specimen performance record from the hover task 14

2. Hover task ratings 15

USING THE BACKWARD TRANSFER PARADIGM TO VALIDATE THE AH-64 SIMULATOR TRAINING RESEARCH ADVANCED TESTBED FOR AVIATION

Introduction

Background

The Army Research Institute Rotary-Wing Aviation Research Unit (RWARU) has designed and developed a unique simulation system called the Simulator Training Research Advanced Testbed for Aviation (STRATA). The primary mission of STRATA is to conduct research to determine the training effectiveness of various simulator and training device configurations. STRATA is a modular system that can be reconfigured to represent a variety of different features of simulators. STRATA, in its present configuration, is a high-fidelity simulation of the AH-64A Apache helicopter. STRATA and its components are described in detail in Kurts and Gainer (1991).

The Backward Transfer Paradigm

One convenient way to assess the training effectiveness of a flight simulator is the backward transfer paradigm (Adams & McAbee, 1961; Kaempf, Cross, & Blackwell, 1989). Highly experienced pilots who are current in the aircraft (but not the simulator) perform aviator tasks from the Aircrew Training Manual (ATM) in the simulator without prior simulator practice. Successful performance of the tasks can be taken as evidence of transfer of training. If backward transfer from aircraft to simulator has been demonstrated, it can be assumed that transfer from simulator to aircraft will also occur.

On the other hand, if experienced pilots perform poorly in the simulator, it can be assumed that cues used in the aircraft are not present in the simulator (Stewart, 1985). In this case, skills possessed by pilots that allow them to perform tasks successfully in the aircraft do not provide them with the capability to perform these tasks in the simulator.

Previous Research

A backward transfer experiment was performed to evaluate the adequacy of the AH-1 Flight and Weapons Simulator (FWS) for the practice of emergency touchdown maneuvers (ETMs) (Kaempf et al., 1989). Performance of ETMs is restricted only to a few training courses, and pilots are prohibited from practicing them as part of routine currency maintenance in their operational units. Only instructor pilots (IPs) would be expected to practice ETMs with enough regularity to be proficient. Subjects were highly experienced IPs who were current in ETM performance as part of

their instructional duties. Each IP was required to pass an aircraft checkride in which ETMs were performed to ATM standards immediately before the simulator practice session. The investigators found backward transfer effects between helicopter and simulator to be extremely weak, as evidenced by a large number of unsatisfactory performance ratings (82%). Fifty-three percent of these resulted in a crash. In post-experimental interviews, participant aviators attributed their performance problems to a lack of visual cues and the poor control input and response characteristics of the FWS. Kaempf et al. (1989) concluded that the AH1FWS could not substitute for the aircraft when practicing ETMs.

The present experiment used a similar rationale. There were a few differences, however. Participants, with one exception, were not AH-64 IPs and had fewer pilot hours in the aircraft than those in the Kaempf et al. (1989) experiment. They were not required to pass an aircraft checkride for specific ATM tasks shortly before performing them in the simulator. The tasks themselves were generally routine aviator tasks, with only one which could be characterized as an emergency procedure (single-engine roll-on landing). Based upon input from AH-64 IPs, any rated AH-64 aviator should be able to perform these ATM tasks in the aircraft.

Overview of Research Approach

Rationale

The experiment was conducted to determine if those skills required to fly the AH-64 transfer to STRATA. The more tasks that can be performed successfully, the greater the degree of backward transfer. Participants were rated as pilot in command (PC) in the AH-64 and had passed a checkride in the aircraft within the past 12 months, in which routine ATM tasks for the aircraft were performed. Each aviator performed the selected ATM tasks in STRATA without the benefit of prior practice. Each task was performed only once; no repetitions were allowed. These tasks were: approach and landing to a confined area, hover taxi, hovering turns, stationary hover, normal takeoff, roll-on landing, rolling takeoff, single-engine roll-on landing, straight and level flight, and terrain flight takeoff.

Objectives

Transfer of training. It was expected that the more proficient the aviator (in terms of total pilot hours in the AH-64) the better the performance in STRATA. Aviators with more hours in the AH-64 should be able to perform the selected tasks in STRATA better than those with fewer hours. Building upon the assumptions presented by Adams and McAbee (1961), an alternative hypothesis would propose that total hours in all aircraft is a

better predictor of performance in the backward transfer experiment than hours in the AH-64. This is predicated upon the assumption that it is generalized aviator skills, not those specific to the particular aircraft, which transfer to the simulator.

Data Recording and Analysis (DRA) System. Another objective of the research was to determine the validity of the performance measures captured by the DRA system. Some of the pre-selected performance measures may correlate with other indicators of performance, (such as IP ratings) while others may not.

Method

Participants

Ten AH-64 aviators, from a Forces Command unit, volunteered to participate in the research (see Table 1). All were males, rated as PCs for the AH-64. None had previously flown STRATA, though all had experience in the AH-64 combat mission simulator (CMS), an operational training simulator whose vision and motion systems are quite different from those of STRATA. The most pertinent differences, in light of the present research, are the vision (CMS has a CRT display; STRATA has a fiber-optic helmet-mounted display, or FOHMD) and motion cuing (CMS has a full motion base; STRATA has a pneumatic G-seat) systems.

Table 1

Participant Background and Experience

Variable	Mean	SD	Minimum	Maximum
Age	31.00	4.37	26.00	41.00
AH-64 Pilot Hours	246.60	171.30	40.00	600.00
AH-64 Copilot Hours	213.00	186.73	0.00	500.00
Total Hr All Aircraft	1179.60	844.73	456.00	2950.00
Total Non AH-64 Hours	710.00	867.54	150.00	2350.00
Days Since Last Flight	5.70	5.52	1.00	15.00
Days Since Checkride	145.50	104.42	15.00	330.00
Days Since Last CMS	27.40	21.68	1.00	60.00
Total CMS Hours	110.00	81.82	25.00	300.00

Procedure

Participant orientation to STRATA. No orientation was given to the participant on any operational features of STRATA having commonality to the aircraft. There was no warm-up or practice session before the experiment began. It was assumed that the aviator rated as PC in the AH-64 should know how to operate a device purporting to simulate the aircraft. However, it was necessary to orient the participant to those features which were unique to STRATA itself, with special attention being given to the FOHMD and the G-seat. Participants were also told to report any problems with the FOHMD, especially any alignment irregularities. They were further told that if they experienced any symptoms of motion sickness or nausea, to report these, and the simulation could be halted at their request.

Mission scenario. The mission scenario was developed with the help of two senior AH-64 IPs, one of whom was a standardization instructor pilot (SIP). SIPs are responsible not only for evaluating student performance, but for assuring that training standards are properly maintained by IPs in the operational units. Each IP had over 1,000 PC hours in the aircraft.

Participants were given a premission briefing on the scenario they were to fly in STRATA. The briefing was given by a retired Army aviator with more than 1,000 hours in the OH-58 helicopter, who also played the role of Air Traffic Controller. They were asked to perform 13 generic ATM tasks, which are listed in all capitals in the scenario summary below. The scenario was held constant for all participants.

The mission began at Falcon Field in Mesa, Arizona, where the pilot would pick up the aircraft to a STATIONARY HOVER, maintaining a heading of 300°, and after 40-50 seconds HOVER TAXI to the departure end of the active runway. Next, he would perform a NORMAL TAKEOFF. After takeoff, the pilot would perform STRAIGHT AND LEVEL FLIGHT at a preassigned altitude, airspeed, and heading, to Phoenix Sky Harbor Airport, approximately 32 km west of Falcon Field. A ROLL-ON LANDING would then be executed on Runway 26L at Sky Harbor. The participant would be asked to set up on the threshold of 26L, pick up to a STATIONARY HOVER and to execute LEFT AND RIGHT HOVERING TURNS. After completing the turns, he would perform a ROLLING TAKEOFF on the same runway. He would then be given an assigned heading, airspeed and altitude for STRAIGHT AND LEVEL FLIGHT toward a Forward Arming and Refueling Point (FARP), located at the base of Red Mountain, approximately 50 km east of Sky Harbor. Upon arrival at the FARP, he would execute a CONFINED AREA APPROACH AND LANDING, then a TERRAIN FLIGHT TAKEOFF with assigned altitude, airspeed, and heading for STRAIGHT AND LEVEL FLIGHT back toward Falcon Field, approximately 13 km southwest of the FARP. En route to Falcon

Field, the left engine would fail unexpectedly, requiring the pilot to execute a SINGLE ENGINE ROLL-ON LANDING at Falcon Field. It was the consensus of both SIPs that the scenario, which took 45 minutes to an hour to perform, was difficult enough to present a challenge to AH-64 pilots.

Dependent Measures

Self Reports and Subjective Measures of Performance

Post-experimental questionnaire. Each participant was asked to complete a questionnaire (Appendix A) at the conclusion of the experiment. The items on the questionnaire were similar to those used by Stewart (1985), and were modified for a helicopter mission setting. Besides routine questions concerning flight experience, the questionnaire consisted of 11 Likert-type items to assess the participant's perception of the degree of similarity between STRATA and the AH-64.

SIP ratings during the experiment. The same SIP who helped developed the scenario also assisted with the evaluation of pilot performance in the experiment. It was not possible to have both AH-64 SIPs present during the entire experiment. One, however, was able to attend all sessions and to perform subjective performance ratings for all participants, using standard ATM criteria. The rating criteria used were: VERY GOOD, GOOD, AVERAGE, MARGINAL, and UNSATISFACTORY.

Automated Performance Measures

DRA performance measures. Examples of representative measures for the ATM tasks are presented in Table 2. The recording of the DRA measures was accomplished through a control program that was triggered by specific events such as location in the visual database, distance from a specific location, airspeed and altitude. For example, during the ROLL-ON LANDING, the DRA would be activated if the aircraft were within a 3 km radius of Sky Harbor Airport. Recording for this task would cease when airspeed dropped below 15 kt. It would resume for the next ATM task, HOVERING TURNS, when the altitude above ground level (AGL) was greater than 0 and airspeed less than 10 kt, at the threshold of Runway 26L. The DRA would automatically turn off when the aircraft was set down again after executing the turns. It would initiate recording once more for the ROLLING TAKEOFF when the aircraft began to roll along the runway beyond the threshold area. For those tasks requiring frequent control inputs (e.g., STATIONARY HOVER), the DRA recorded at approximately 9 Hz. For other tasks, such as STRAIGHT AND LEVEL FLIGHT en route, recording frequency was 1 or 2 Hz.

Table 2

Performance Measures for ATM Tasks

Task(s) (Freq)	Performance Measures
Hover and Hover Taxi (9 Hz) (Combined)	Altitude above ground level (AGL), Airspeed, Heading, Turn Rate, Lateral Cyclic Displacement, Pedal Displacement.
Normal Takeoff (2 Hz)	Altitude AGL/mean sea level (MSL), Airspeed, Heading, Rate of Climb, Distance from Falcon Field, Roll, Pitch, Turn Rate.
Straight & Level Flight (1 Hz) (Repeated 3X)	Altitude AGL/MSL, Airspeed, Heading, Rate of Climb, Distance from Destination, Roll, Pitch, Turn Rate.
Roll-on Landing (1 Hz)	Altitude AGL/MSL, Airspeed, Heading, Rate of Climb, Distance from Sky Harbor, Roll, Pitch, Turn Rate.
Hover and Hovering Turns (9 Hz) (Combined)	Altitude AGL, Airspeed, Heading, Turn Rate, Pitch, Lateral Cyclic Displacement, Pedal Displacement, Engine Torque.
Rolling Takeoff (2 Hz)	Altitude AGL/MSL, Airspeed, Heading, Rate of Climb, Distance from Sky Harbor, Pitch, Roll.
Confined Area Landing (1 Hz)	Altitude AGL/MSL, Airspeed, Heading, Rate of Climb, Distance from Destination, Pitch, Roll, Turn Rate.
Terrain Flight Takeoff (2 Hz)	Altitude AGL/MSL, Airspeed, Heading, Rate of Climb, Pitch, Roll, Turn Rate.
Single Engine Roll-on Landing (1 Hz)	Altitude AGL/MSL, Airspeed, Heading, Rate of Climb, Pitch, Roll, Turn Rate, Distance from Falcon Field, Collective Position, Lateral/ Longitudinal Cyclic Position, Pedal Position, Engine Torque.

Note. Some of the tasks, because they were part of the same event during a given mission segment, were combined. Another task, straight and level flight, was repeated as the aircraft flew between waypoints in the scenario.

Performance rankings of DRA output after the experiment. The subjective real-time performance ratings were given by a single SIP. It was not possible to employ a paired set of independent real-time ratings from two or more IPs during the experiment. For this reason, the reliability of a single set of ratings may be called to question. To address this question, a post-experimental rank-ordering of performance on the initial hover task, based upon DRA output, was conducted after the experiment for purposes of concurrent validation.

The hover task was chosen because (a) subject matter experts considered it to be one of the more difficult ATM tasks for the aircraft, (b) it was the first task performed and hence a relatively "pure" measure of backward transfer, and (c) ATM performance standards for the task are set forth more explicitly than for some other tasks.

There were four judges. Three IPs and one retired Army aviator were asked to make independent ratings and rank orderings of performance, using graphical output from the DRA as stimulus materials, for the STATIONARY HOVER task. Each participant was identified only by letter (A through J, randomized). Thus judges were blind to participants' identities and had only the performance measures (airspeed, heading, altitude, and lateral cyclic displacement) to use as criteria. Two judges were the same IPs who had served as consultants during the experiment. One of these two was the SIP who had administered the real-time performance ratings. The third IP was newly assigned to the STRATA project and had not participated in the experiment. The fourth judge, the retired aviator, was currently employed by the simulator manufacturer and had assisted with the conduct of the experiment, but had not participated in the ratings during its course.

Results

Participant Evaluation of STRATA

Structured questionnaire responses. After the simulator session, each participant was asked to indicate the degree to which he perceived STRATA's flight characteristics to be similar or dissimilar (6-point scale) to the aircraft. Rating alternatives were: very different/different/somewhat different/somewhat similar/similar/very similar. The scales were keyed positively so that the higher the rating, the higher the degree of perceived similarity to the aircraft.

Table 3, below, shows that most participants perceived the simulator's flight characteristics to be similar to those of the aircraft.

Table 3

Participant Ratings of Similarity of STRATA to Aircraft

Item	Mean	SD	Minimum	Maximum
1 General	4.80	.63	4.00	6.00
2 Pitch	4.50	1.08	2.00	6.00
3 Roll	4.60	.52	4.00	5.00
4 Yaw	4.30	1.25	2.00	6.00
5 Acceleration	5.00	.67	4.00	6.00
6 Cyclic	5.40	.84	4.00	6.00
7 Collective	4.60	1.17	3.00	6.00
8 Hover	4.40	1.08	3.00	6.00
9 Pedals	5.30	.67	4.00	6.00
10 Turns	5.00	.67	4.00	6.00
11 Power	4.80	1.23	2.00	6.00

Highest similarity ratings concerned lateral control characteristics. The positivity of these ratings is noteworthy when we recall that STRATA has no full motion base.

Open-ended comments. Participants were invited to provide open-ended, spontaneous comments on their impressions of STRATA, to the degree that they found its performance like or unlike the AH-64. All provided some comments. The most frequent comments (eight mentions) concerned the lack of adequate visual cues, such as texture or contrast, for hovering and low-level flight.

The next most frequent category of comments was concerned with positive reactions to the simulation in general and specific references to how STRATA handled like the AH-64 (six mentions).

Participants tended to be more ambivalent about motion cues in general and the G-seat in particular. Five mentioned that G-seat motion cues were frequently dissimilar to those in the aircraft. A listing of all open-ended comments appears in Appendix B.

Self-reports of motion sickness. No participants reported any adverse symptoms of nausea or motion sickness during or immediately after the experiment.

SIP Ratings of Performance

Summary of ratings. Table 4 shows the frequency distribution of ratings given by an AH-64 SIP during the experiment, for each of the ATM tasks performed. The table also presents the distribution of these performance ratings across all

subjects. Of the 130 task events (10 participants each performing 13 tasks), 88.5% were performed satisfactorily, while the remaining 11.5% were rated as indicating unsatisfactory performance (one was the result of a crash). Of the 130 task events, 24 (18.5%) were classed as marginally satisfactory (marginally satisfying the ATM standard for the task). The remaining 70% of the task events were classified as clearly satisfactory, ranging from average to very good. Eight participants showed unsatisfactory performance on at least one task.

Examining the marginal means for each task, it is clear that performance was worst for the confined area landing, and best for the single-engine roll-on landing. An obvious question is whether or not there was a general trend for rated performance to improve across tasks (and across time) as the simulation progressed. A one-way ANOVA showed no significant trend ($\underline{F}$ = 1.27, $\underline{df}$ = 12/117, $\underline{p}$<.25).

Comparison between tasks, even similar ones, is difficult because of topographical variations within the visual database comprising the scenario. There are three instances of straight and level flight. The first instance takes place over the Phoenix metropolitan area, which is situated in the flat floor of a valley. The other two instances occur over rugged, mountainous terrain northeast of Phoenix where elevation is variable. Some participants complained about using altitude above ground level (AGL) as a criterion because they found themselves "chasing the radar altimeter." It was decided to stay with AGL, since this presented a more rigorous test of what pilots can accomplish in the STRATA simulator.

Table 4

Frequency Distribution of Performance Ratings for 13 Tasks

Task Description	Performance Rating[a]					
	VG	G	A	M	U	Mean
1. Hover (Falcon)	2	2	1	2	3	2.8
2. Hover taxi	2	4	0	2	2	3.2
3. Normal takeoff	0	8	1	1	0	3.7
4. Straight & level flight	4	3	1	1	1	3.8
5. Roll-on landing	2	4	1	2	1	3.4
6. Hover (Sky Harbor)	0	6	1	1	2	3.1
7. Hovering turns	2	1	2	3	2	2.8
8 Rolling takeoff	1	2	4	2	1	3.0
9. Straight & level flight	3	3	1	3	0	3.6
10. Confined area landing	0	2	3	3	2	2.5
11. Terrain flight takeoff	2	5	1	2	0	3.7
12. Straight & level flight	0	7	1	1	1	3.4
13. Single-engine landing	3	6	0	1	0	4.1
Totals	21	53	17	24	15	3.3

[a]VG = very good (5); G = good (4); A = average (3); M = marginally satisfactory (2); U = unsatisfactory (1).

<u>Performance and AH-64 PC hours</u>. The range of PC hours was truncated. This may in part be due to current Department of Defense restrictions on flying hours. The distribution of self-reported PC hours showed two values tied at the median (200). The next highest was 300. Thus a simple median split was not practical. Two categories were formed by placing those values of 300 and above into the high time category, and 200 and below into the low time category. A comparison of the distribution of ratings between this subsample and those with fewer than 300 hours would nevertheless provide adequate expected cell frequencies for a χ^2 test. Comparing these two distributions (see Table 5 below), yielded a χ^2 of 16.12, which at four degrees of freedom was significant beyond the .003 level. Those participants with over 300 PC hours had a higher percentage of very good ratings (29% vs. 8%) as well as lower percentages of

Marginal ratings (10% vs. 24%) and UNSATS (6% vs. 15%). Thus, the hypothesis that those pilots with more PC hours would outperform those with fewer PC hours was confirmed.

Table 5

Frequency Distributions of Performance Ratings as a Function of Flight Experience (Percentages in Parentheses)

PC Hours	Very Good	Good	Average	Marginal	UNSAT
High (n = 4)	15 (29)	20 (39)	9 (17)	5 (9)	3 (6)
Low (n = 6)	6 (8)	33 (42)	8 (11)	19 (24)	12 (15)
Copilot Hours					
High (n = 5)	9 (14)	26 (40)	2 (3)	17 (26)	11 (17)
Low (n = 5)	12 (18)	27 (42)	15 (23)	7 (11)	4 (6)
Total Hours in All Aircraft Types					
High (n = 5)	12 (18)	29 (45)	13 (20)	3 (5)	8 (12)
Low (n = 5)	9 (14)	24 (37)	4 (6)	21 (32)	7 (11)

AH-64 copilot (CP) hours. Typically, an AH-64 pilot must spend time in the front seat as a copilot before moving up to PC. The number of self-reported CP hours ranged from 0 to 500. Unlike PC hours, there was a definite split at the median, from 110 to 300 hours. Thus it was practical to divide the sample into two subgroups. Table 5 also displays the rating distributions by AH-64 CP hours. The trend was the opposite from that found for PC hours. The low-time aviators had a lower percentage of task events rated as Marginal than did the high-time aviators (11% vs. 26%). Likewise, 17% of the task events were rated as UNSAT for those pilots with high CP hours, vs. 6% for those with low CP hours. For those with high CP hours, 57% of all ratings were for average or better performance; for those with low hours, the respective percentage was 83. The association between CP hours and rating distribution was significant ($X^2 = 17.96$, $df = 4$, $p<.001$). For this particular experiment, the fewer the CP hours, the better the performance on the ATM tasks. This may seem counterintuitive at first. However, it makes sense when we realize that AH-64 copilots are primarily responsible for operating the weapons systems of the aircraft, rather than flying it. Because of the high workload situation imposed by these duties, there is little opportunity to fly the aircraft. Thus flying skills may deteriorate. An

alternative, though by no means mutually exclusive explanation, for these findings would be that the more proficient copilots move up to the back seat faster, and hence remain copilots for less time.

Total flight hours, all aircraft. Total flight hours ranged from 456 to 2950. The ratings of all 10 participants were split at the median (872.5 hr). The rating distributions of these two subsamples also appear in Table 5. A χ^2 of 19.32 was significant for four degrees of freedom (p<.001). For the more experienced pilots, 83% of ratings on 65 task events were for average performance or better, 5% were Marginal, and 12% UNSAT. For the less experienced pilots, 57% of the task events were performed at a level of average or above performance. Thirty-two percent were rated Marginal, and 11% UNSAT. It would seem that the difference in performance between the two subsamples was primarily due to differences in the incidence of Marginal performance.

Marginal and unsatisfactory performance. The results of the preceding analyses suggested that ratings on ATM tasks for which performance was judged to be Marginally satisfactory or UNSAT may be sensitive metrics of performance. Only one participant had no UNSATS or Marginals. This pilot also had the most PC hours (600), and was a close second in total hours in other aircraft (2,300). The number of UNSATS ranged from 0 to 4 with a mean of 1.60. Two participants had none. Marginals ranged from 0 to 6 with a mean of 2.5. Four participants had none. The self-reported pilot hours in the AH-64, total flight hours in all aircraft, and time elapsed since last flight and checkride were correlated with the number of UNSATS and Marginals.

Table 6 shows the intercorrelations between these measures. An examination of Table 6 shows that Marginals were negatively and significantly correlated with total flight hours in all aircraft types. The total number of AH-64 pilot hours, though in the expected direction, does not correlate significantly with the number of Marginals or UNSATS. Thus total flight hours in all aircraft seems to have been the strongest predictor of the number of Marginals, but not UNSATS. Another correlation that was found to approach significance in the expected direction was the time since the last AH-64 flight and the number of Marginals. The greater the elapsed time, the greater the number of Marginals.

The correlation between the time since last checkride and the number of Marginals (-.83, p<.005) seems counterintuitive at first glance. However, time since last checkride is also highly correlated with total hours in all aircraft (.89, p<.005). It would seem then, that the more experienced the pilot, the more time that had passed since the last checkride. Another interesting finding was the significant positive correlation between copilot hours and Marginals. The reason for this correlation is somewhat unclear, though two possible explanations

for the strong association between CP hours and poorer performance have been offered. Although not appearing in the table, it is interesting to note that the number of hours spent in the AH64CMS correlated highly with only one variable: CP hours (r = .66, p<.05). This is consistent with the previously-noted high negative rs between CP hours and pilot performance measures. It seems that CMS time substitutes for actual aircraft time for AH-64 copilots, who spend most of their time managing the offensive weapons systems of the aircraft.

Table 6

Pearson Intercorrelations: Flight Hours, Time Since Last Flight, and Checkride vs. UNSATS and Marginals

Variable	PLTHR	CPHR	TOTHR	LFLT	LCHR	UNSAT
PLTHR	1.00					
CPHR	-.40	1.00				
TOTHR	.42	-.35	1.00			
LFLT	-.29	.00	-.47	1.00		
LCHR	.32	-.35	.89c	-.50	1.00	
UNSAT	-.51	.69b	-.43	.12	-.37	1.00
MARG	-.32	.18	-.67b	.61a	-.83c	.37

Note. PLTHR = PC hours; CPHR = CP hours; TOTHR = total hours, all aircraft; LFLT = time since last flight; LCHR = time since last checkride; UNSAT = unsatisfactory; MARG = marginal. a = <.07; b = <.05; c = <.01 (all p's two-tailed)

SIP ratings and their correlation with post-experimental rankings on the hover task. It is difficult to assess the reliability of the DRA measures without an independent criterion. For this reason, an exercise was planned in which ratings by one SIP made during the experiment would be compared to post-experimental rank orderings of performance on a selected task, based solely on DRA output. The DRA output of selected measures on the initial stationary hover task was used for this exercise.

The performance objectives for the hover task are clearly defined. In order to meet ATM standards, the pilot must maintain a constant heading within + or -10°, an altitude of 5 feet

(within a range of +-2 ft), and should not allow the aircraft to drift more than 3 feet. Besides looking for performance along these dimensions, the IP also considers control input, especially lateral movement of the cyclic pitch control. An experienced pilot proficient at hovering should not overcontrol the aircraft by making excessive cyclic inputs.

Recall that ratings and rank orderings of the hover task were performed by three IPs and one retired Army aviator. DRA measures of altitude, airspeed, heading, and lateral cyclic displacement, plotted against time in seconds, provided data for the ratings of the 10 aviators. All data were presented as line graphs, with time in seconds on the abscissa, and the performance measures on the ordinate. Figure 1 shows a specimen record for one participant whose performance was rated as "very good" during the experiment, and by all four judges afterward. Note that hover height seems greater than formal ATM standards. Interviews with participants and with other AH-64 pilots indicated a preference for hovering at altitudes of approximately 10 feet. At lower altitudes, turbulence causes discomfort.

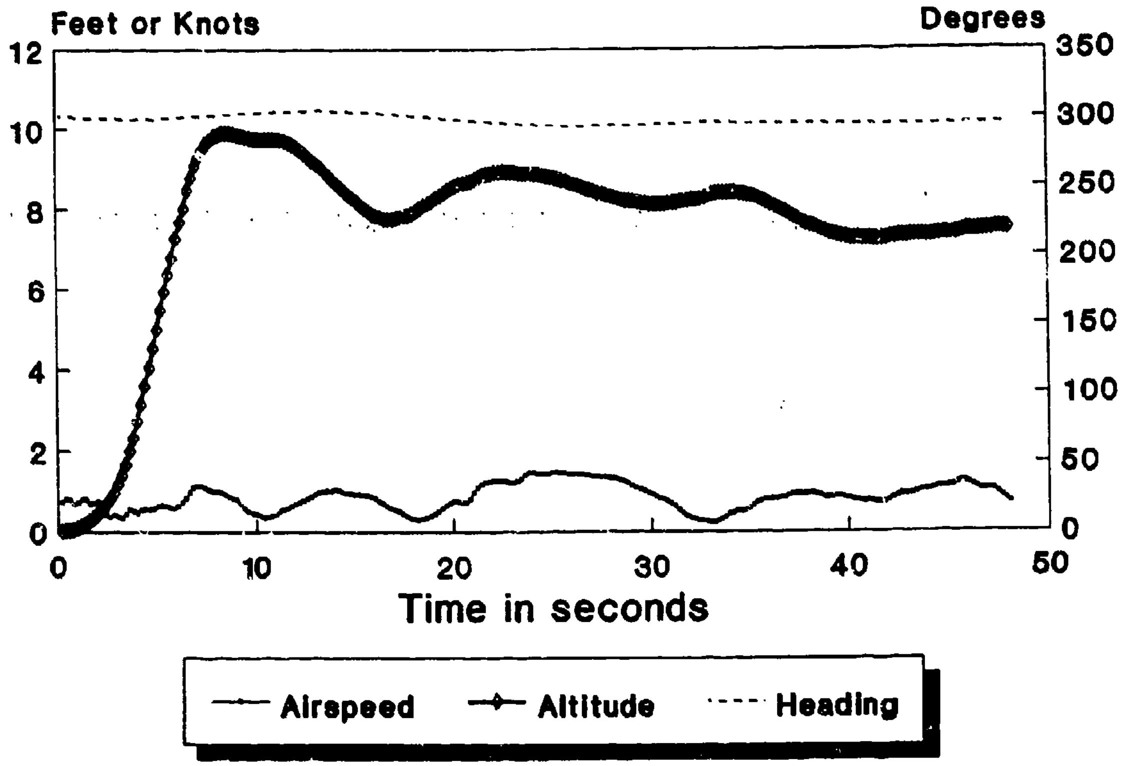
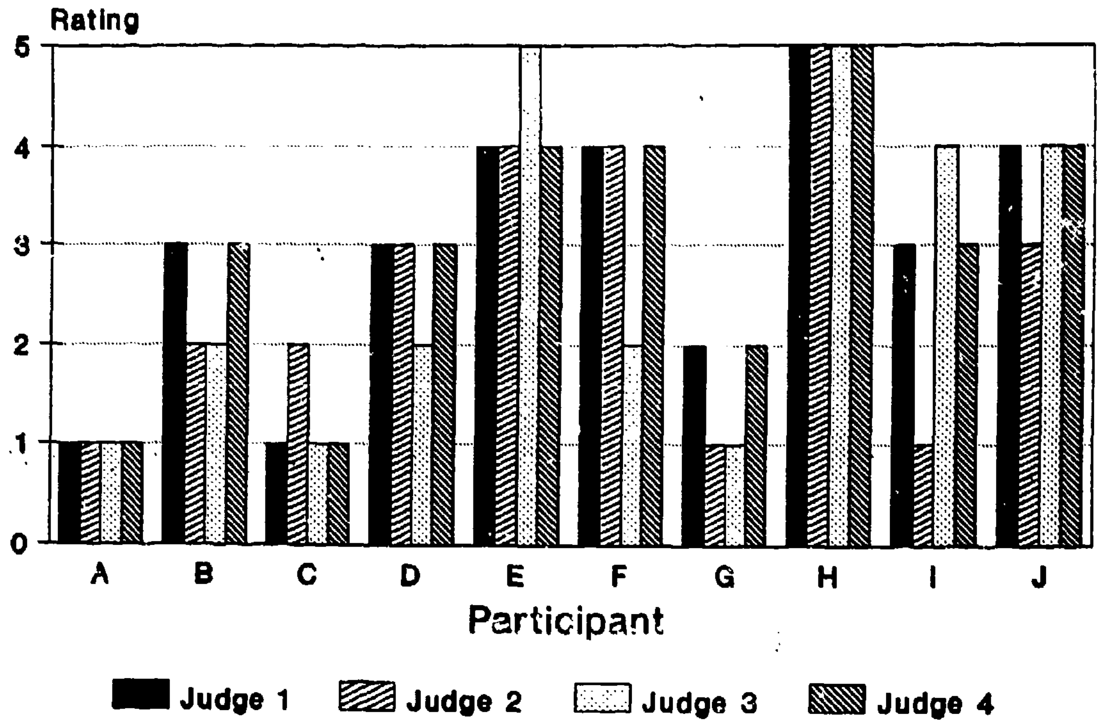


Figure 1. Specimen performance record from the hover task (pilot rated as "very good").

Table 7 presents ratings and rank orderings for the hover task. Participants were ranked from 1 = best to 10 = worst. The degree of concordance and average Spearman r_s were highly significant, indicating a high degree of inter-judge reliability. Judge 2 is the SIP who provided subjective ratings of pilot performance during the experiment. His hover task ratings made during the experiment were compared to the average post experimental rankings assigned by the other three raters. A correlation of -.77 ($p<.02$) indicated that performance ratings of the stationary hover during the experiment had moderate concurrent validity. Ratings are graphically presented in Figure 2, below.



Note. 1 = Unsatisfactory; 5 = Very Good.

Figure 2. Hover task ratings.

Table 7

Ratings and Rank-Orderings of 10 Participants on Task 1: Pick Up to Hover (Ranks are in Parentheses)

Participant	Judge 1 (Rank)	Judge 2 (Rank)	Judge 3 (Rank)	Judge 4 (Rank)
A	UNSAT (9)	UNSAT (9)	UNSAT (9)	UNSAT (9)
B	AVERAGE (5)	MARGINAL (6)	MARGINAL (5)	AVERAGE (5)
C	UNSAT (10)	MARGINAL (7)	UNSAT (8)	UNSAT (10)
D	AVERAGE (7)	AVERAGE (5)	MARGINAL (7)	AVERAGE (7)
E	GOOD (2)	GOOD (2)	VERY GOOD (2)	GOOD (2)
F	GOOD (3)	GOOD (3)	MARGINAL (6)	GOOD (3)
G	MARGINAL (8)	UNSAT (10)	UNSAT (10)	MARGINAL (8)
H	VERY GOOD (1)	VERY GOOD (1)	VERY GOOD (1)	VERY GOOD (1)
I	AVERAGE (6)	UNSAT (8)	GOOD (4)	AVERAGE (6)
J	GOOD (4)	AVERAGE (4)	GOOD (3)	GOOD (4)
Overall Ratings	M = 2.90 S = 1.37	M = 2.60 S = 1.43	M = 2.80 S = 1.60	M = 3.00 S = 1.33

Note. Kendall's $W = .91$, $p<.004$; Average $r_s = .88$ ($p<.005$)

DRA Measures and Their Correlations

The DRA measures sampled at the rate of approximately 9 Hz for the hover task are presented in Table 8. The standard error of the mean was used as a candidate performance measure because 450 observations were taken on each participant during the hover task. Thus each participant had a standard error score for the variable being measured. The mean standard error for each variable is similar to a grand mean for all participants.

This could possibly provide an index of steadiness and variability on a task which requires that parameters such as speed and altitude be kept constant. For example, if we examine Heading, we can see that the standard error ranged from a low of .12 for one participant to a high of 3.42. The intercorrelations of these measures and the subjective performance ratings appear in Table 8. Because the standard error airspeed and lateral cyclic displacement data showed a high degree of variation and were highly skewed, all performance measures were converted to rank-order data. For purposes of maintaining consistency with other evaluative measures and avoiding confusion, the rank-order data were coded so that a high number corresponded to a high ranking.

Table 8

Performance Measures (Means and Standard Errors) for Initial Hover Task

Variable	Mean	SD	Minimum	Maximum
Mean Airspeed (Kt.)	2.00	1.28	.79	5.25
SE Airspeed	.06	.04	.03	.18
Mean Heading°	309.14	11.45	295.51	332.17
SE Heading	.73	.97	.12	3.42
Mean Altitude	5.68	5.02	.60	17.49
SE Altitude	.14	.20	.03	.69
SE Lateral Cyclic	.03	.05	.00	.17

Table 9

Spearman Rank Order Correlations of DRA and Other Performance Measures for Initial Hover Task

Variable	MAS	SAS	MHDG	SHDG	MALT	SALT	PLTHR	RAT	RNK
MAS	1.00								
SAS	.73	1.00							
MHDG	.88	.86	1.00						
SHDG	.47	.58	.55	1.00					
MALT	-.06	-.01	.19	.37	1.00				
SALT	.16	.02	.27	.39	.92	1.00			
PLTHR	-.16	.46	.15	.27	.43	.20	1.00		
RAT	-.87	-.46	-.68	-.40	.14	.02	.40	1.00	
RNK	-.83	-.66	-.79	-.73	-.18	-.36	.02	-.77	1.00
SCYC	.78	.45	.60	.37	-.23	.02	-.38	-.79	-.56

Note. There were 450 repeated measures per participant over a time period of approximately 50 seconds. These data were collapsed across this time period. MAS = Mean airspeed; SAS = Standard error airspeed; MHDG = Mean heading; MALT = Mean altitude; SALT = Standard error altitude; PLTHR = PC hours; RAT = Performance rating during experiment; RNK = Post experimental ranking by judges 1,3,4; SCYC = Standard error, lateral cyclic displacement. For $p<.05$ (two-tail), r (critical) = .63.

An examination of Table 8 reveals that several performance measures were significantly correlated with the subjective ratings of performance given during the experiment. These were: standard error lateral cyclic displacement (SCYC), mean airspeed (MAS), and mean heading (MHDG). The average rankings given after the experiment by the other three raters correlated significantly with standard error airspeed (SAS), standard error heading (SHDG), MAS, and MHDG. These are performance variables closely related to the formal ATM standards for hovering. For example, the lower and less variable the airspeed and less the variation in heading, the better should be the subject's performance. The highest correlation for both rating situations was for MAS. This is not surprising, since AS is an obvious and easily observable criterion for the task. Although the ATM sets explicit standards for hover altitude, MALT was neither correlated with the ratings or post-experimental rankings. It is interesting as well to note that, contrary to expectations, the total number of pilot hours (PLHR) in the AH-64 correlated significantly with neither the DRA measures, nor the subjective performance ratings/rankings. Although not shown in Table 8, it is noteworthy that the total score, based on the SIP ratings, summed across all 13 tasks, did correlate significantly with PLTHR (r = .67, $p<.05$), but not with any of the other variables.

Discussion

Participant Evaluations of the Simulation

STRATA was rated highly by participants in terms of perceived handling similarity to the AH-64. Most participants reported that it generally handled like the aircraft. Of the specific handling characteristics, acceleration-deceleration, cyclic response, turns, and pedal response received the highest mean ratings. Open-ended questions revealed that they were much less impressed with the visual display system. Most stated that it lacked the resolution and contrast necessary for effective performance of low-level tasks like hovering.

Demonstration of Backward Transfer

Both the subjective and DRA data indicated that backward transfer was successfully demonstrated in STRATA. Pilots who were current in the AH-64 helicopter were able to complete a simulated mission scenario, with no preflight orientation or warm-up. This was true for both high-time and low-time pilots. There were few instances of unsatisfactory performance, only one of which involved a crash. This stands out in contrast to Kaempf et al., (1989), where the percentage of unsatisfactory performances approximated the number of successful ones in the present research. Also, more than half of the unsatisfactory performances in the former experiment involved a crash.

Subjective and Objective Ratings and Rankings

Correlations between DRA and SIP real-time ratings for the hover task generally supported the conclusion that DRA measures for this task are valid and potentially useful as evaluative criteria. The DRA also served as a benchmark for validating the subjective real-time ratings of performance on the hover task.

Post-experimental rating and ranking exercises could also provide insight into the relative weights that IPs assign to different performance parameters, and whether these are stable individual differences, or situationally determined. Although concordance between raters was high, there was one instance (Participant I) where the ratings were quite discordant. It is cases like this that may be very interesting if our goal is determining what internal anchors and criteria IPs use when rating pilot performance.

Limitations

The sample size was small for the interpretation of correlations between self-report questionnaire data and performance evaluations. Still, some correlations were quite intriguing and would seem to warrant replication of the

experiment in order to increase sample size. A very interesting finding was the high negative correlation between the number of marginal performance ratings and the total number of flight hours in all aircraft types. This implies that general flying skills may be a more important determinant of performance in STRATA than those specific to the AH-64. The negative correlation between pilot hours in the AH-64 and the number of instances of unsatisfactory performance approaches significance. For a larger sample, these two correlations, if stable, would raise some intriguing questions about transfer of training.

Suggestions for Future Research

Adams and McAbee (1961) noted that it may not be wise to confine backward transfer experiments only to the most proficient aviators. The level of skill integration and the manner in which cues are processed may be quite different when this group is compared to aviators of lesser experience. Thus highly-proficient aviators may possess more generalized skills than novice aviators. Consequently, it may be the general skills of the highly-proficient aviators, and not their experience in the AH-64, that transfer to the simulator. For this reason Adams and McAbee suggested that the backward transfer paradigm was an excellent medium for studying skills integration in pilots differing in experience. So far this capability has not been exploited.

This presents a cogent argument for employing subjects differing widely in experience. There are several interesting hypotheses that could be derived from this type of sample. We do not know, for example, if it is pilot hours in the specific aircraft or total pilot hours, regardless of the aircraft, that is the strongest predictor of performance in the simulator. Moreover, little is known about the relative dependence of different tasks on general and specific skills. Therefore, it would be of interest to note whether it is AH-64 pilot hours or total flight hours that best predict performance in STRATA.

The question as to whether specific or general piloting skills transfer from simulator to aircraft or vice versa may be overly simplistic. Both should transfer, but the degree of transfer could depend upon the performance requirements of the task. Some tasks may be more specialized than others. For example, it would seem reasonable to suppose that pilots with many hours in several aircraft types would have acquired general air sense and adaptive skills which should allow them to perform adequately across a broad range of tasks. Thus, few of their performances should be marginal. On the other hand, some piloting tasks, such as hovering turns, may be more dependent upon familiarity with the particular aircraft. Examples would be

tasks which require a high degree of familiarity with the AH-64's control loading, which is different from that of many other helicopters.

It is important to examine the implications of the STRATA backward transfer results in the context of in-simulator transfer of training experimentation. This, after all, is a foundation for the testbed approach to simulator and training system development. If training devices of varying complexity can be derived from STRATA, we must ask if STRATA, in its current configuration, could be used as a criterion for evaluating their performance. The backward transfer results suggest that it could, but this statement must be tempered with caution. First, the validation of STRATA involved a small number of ATM tasks. Secondly, the sample, as has been previously stated, was small. Consequently, it would be seem that before STRATA can be employed routinely as a criterion, the current research should be expanded to include more pilots and more ATM tasks. For future validation research, it would also be desirable to require pilots to pass a checkride in the AH-64 immediately prior to the experiment, or as close in time to the experiment as possible.

Another research approach would be to employ STRATA in a forward transfer of training paradigm, using pilots who were transitioning from initial entry rotary wing training to the aviator qualification course in the AH-64. This research could provide guidance as to what proportion of training time could be allocated to simulator and to aircraft. Such information would be especially valuable at a time when the Army is re-examining its use of simulation in the training and sustainment of flying skills.

References

Adams, J. A., & McAbee, W. H. (1961). A program for a functional evaluation of the GAM-83 Melpar Trainer (Report No. APGC-TN-61-41). Eglin AFB, FL: U.S. Air Force Air Proving Grounds. (AD 268220)

Kaempf, G. L., Cross, K. D., & Blackwell, N. J. (1989). Backward transfer and skill acquisition in the AH-1 flight and weapons simulator (ARI Research Report 1537). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A213 432)

Kurts, D., & Gainer, C. A. (1991). The use of a dedicated test-bed to evaluate simulator training effectiveness. In Piloted Simulation Effectiveness Conference, AGARD Conference Proceedings No. 513 (pp. 11-1 to 11-9). Brussels, Belgium: North Atlantic Treaty Organization.

Stewart, J. E. (1985). Learning and performance in an air refueling part-task trainer: A preliminary data analysis. Proceedings of the Human Factors Society 29th Annual Meeting (pp. 408-411). Santa Monica, CA: Human Factors Society.

APPENDIX A

Participant Questionnaire

Backward Transfer of Training Experiment

PARTICIPANT NUMBER_____

Backward Transfer of Training Experiment

BACKWARD TRANSFER simply means the degree to which aviators who are already proficient in an aircraft type are able to perform in a simulator which is supposed to represent the aircraft. The more faithfully the simulator models the aircraft, then the better rated aviators should perform in it.

For this experiment, you will be asked to "fly" a simulator representing the AH-64. The simulator is called STRATA (Simulator Training Research Advanced Testbed for Aviation). We will ask you to imagine that you are going to fly the actual aircraft. There will be no orientation or warm-up. You will, however, be given informal orientation on those characteristics that are unique to the STRATA (e.g., the Fiber Optic Helmet Mounted Display).

PART I: BACKGROUND QUESTIONNAIRE

There are a few questions that we would like to ask, for data analysis only, before we begin. This is ANONYMOUS, and there is no way that your name, SSN and other identifying characteristics can be determined. We have simply assigned you a number corresponding to the order in which you performed the experiment.

1. How many pilot hours have you had in the AH-64?________________hours

2. How many copilot hours have you had in the AH-64?________________hours

3. What is the APPROXIMATE date of your last flight in the AH-64? ________.

4. How long has it been since your last CHECKRIDE in the AH-64? ________months.

5. Indicate below the approximate hours you have had in other aircraft, including fixed-wing.

Aircraft	APPROXIMATE Hours	APPROXIMATE Date of Last Flight

6. What is your age, rounded to the nearest year?_____

7. What is your current rank?______

8. Approximately how may hours have you had in the AH-64 Combat Mission Simulator (CMS)? __________hours. About how long has it been since your last CMS session?____________.

ART II: EVALUATION OF SIMULATION EXERCISE

Please complete this section AFTER you have performed the simulated flight scenario. We are interested in the degree to which STRATA models the performance of the actual aircraft. Your responses to the following questions would be of great value to us. Please indicate your impressions by placing and **X** in the appropriate box below:

1. IN GENERAL, how SIMILAR were the flight characteristics of STRATA to those of the AH-64?

Very Different	Different	Somewhat Different	Somewhat Similar	Similar	Very Similar

IN PARTICULAR: How would you judge the SIMILARITY of the following performance characteristics of STRATA to those of the AH-64 ?

2. Control about the PITCH axis.

Very Different	Different	Somewhat Different	Somewhat Similar	Similar	Very Similar

3. Control about the ROLL axis.

Very Different	Different	Somewhat Different	Somewhat Similar	Similar	Very Similar

4. Control about the YAW axis.

Very Different	Different	Somewhat Different	Somewhat Similar	Similar	Very Similar

5. ACCELERATION and DECELERATION.

Very Different	Different	Somewhat Different	Somewhat Similar	Similar	Very Similar

6. Responsiveness to CYCLIC Inputs.

Very Different	Different	Somewhat Different	Somewhat Similar	Similar	Very Similar

7. Responsiveness to COLLECTIVE Inputs.

Very Different	Different	Somewhat Different	Somewhat Similar	Similar	Very Similar

8. Performance during HOVERING.

Very Different	Different	Somewhat Different	Somewhat Similar	Similar	Very Similar

9. Responsiveness to PEDAL INPUTS.

Very Different	Different	Somewhat Different	Somewhat Similar	Similar	Very Similar

10. Performance during TURNS.

Very Different	Different	Somewhat Different	Somewhat Similar	Similar	Very Similar

11. Performance during POWER CHANGES.

Very Different	Different	Somewhat Different	Somewhat Similar	Similar	Very Similar

12. We would be interested in any additional impressions that you may have of the simulation in which you have just participated. We are especially interested in the ways that you found STRATA to be LIKE and UNLIKE the AH-64. If you wish, you can write your impressions below. If you need more space, you can continue on the blank sheet provided.

APPENDIX B

Open-Ended Comments From Questionnaire

Open-Ended Comments from Questionnaire.

Participant	Comments
1.	I found it lacking in some contrast that a pilot would need to aid in hovering. I used symbology often to perform hovering turns, which I normally would not (do) in daytime flight (real or CMS). I also noticed some flickering in the picture. After a long flight period this could become very fatiguing.
2.	Flight control feel was very good. I have trouble hovering any simulation device and I believe that this is due to lack of depth perception and visual references. I felt a little bit bound by the fiber optic cables. Helmet was very tight.
3.	When picking STRATA up to a hover I found it hard to use visual cues for altitude reference. I had to resort to instrument cues. Azimuth and drift were not as hard although I found myself drifting due to using radar altimeter for height AGL. Also, clarity of the instruments and gauges was a little distorted which made it harder to use a good cross check. This meant a larger amount of time with my head in the cockpit. When picking up to a hover and a few other instances (hovering mainly), I felt the seat was an accurate duplication of the feeling of the actual AH-64. I'm not sure exactly how to explain the difference. I think mainly in the way the seat inflated. Normal or straight-and-level flight was hard to maintain due to visual references. The helmet was somewhat restrictive when turning left or right at any great distance. When bringing the aircraft up so it's light on the wheels for rolling takeoff or ready for normal takeoff there was really no sensation of being light until it was off the ground. Had to use a lot of instrument reference.

Participant	Comments
4.	I would have liked to have a HDU so that I could have flight symbology while looking outside the cockpit. I felt that the graphics were not giving the visual cues that I needed to do hovering turns properly. Left optics were off center four different times; stepping occurred once. Picture built and took away the mountain at the FARP. I continually seemed to be chasing the power setting to maintain airspeed and altitude. I seemed to drift on the active for a second or two before I noticed it. I found myself mentally blocking out the cues from the G-seat because the cues seemed to be irritating and bothersome. Helmet was too tight around the ears. Picture seemed to be gritty or dirty. Pedal adjustment was wrong.
5.	The tail wheel lock I couldn't get to work. It worked but I couldn't tell if it was unlocked or I was dragging my tail wheel. Power requirements for rolling takeoff: I came up with less power than normal. I really liked the simulator. I would like a better way to feel the motion of the aircraft. Pedals would not extend out far enough.
6.	It seemed as if the aircraft would hover taxi a little faster than the actual aircraft. The graphics were adequate for this but if a little more texture was added you have more of a sense of motion. The force trim interrupt didn't seem like it would hold its new position which made hovering a little difficult. Collective friction is needed. I did not like the feeling of the seat. It was good for doing high/low-G maneuvers but it tended to confuse me more while hovering.
7.	I found that after a few minutes it flew very much like the real aircraft; however, I have a few observations: a. Force trim feels a little different. I found it hard to get the aircraft trimmed up at a hover. It was a little better in flight, however. b. I had problems feeling the aircraft touch down. I never really knew when it was on the ground without cross checking my instruments. c. The pedals are much too close. It causes my yaw inputs to be over emphasized, especially at hover.

Participant	Comments
	d. I believe the device is very good; however, we could have a little more detail.
	e. I found the helmet after a period of time to be restrictive, because of the computer attachments, when I went to look left or right, especially during hovering turns.
8.	The inset was tilted to the right which gave me the impression I was in a constant turn. The attitude indicator did not function the first part of the flight. The VDU was fading in and out. I could not always see the heading tape. The G-seat fell off line. But over all, I think that after a few practice flights a person could fly well. The quality of the visuals is very low, and the G-seat does not react to small collective or pitch changes to the extent that the pilot can feel them.
9.	I used the VDU to see objects at a distance. I would like to see the horizon line sharper. The aircraft floats during ground taxi. Pedal inputs on the ground need improving. Adjust the collective friction; it feels too light or too heavy. Pitch is too sensitive. Roll is not sensitive enough. The cockpit feels like an AH-64. A great experience! Not enough visual close by for cuing. Sharper resolution is needed for very low flying.
10.	I found it to be a very good simulation in most every way. However, the depth perception was difficult trying to hover over the runway with very little references.